A game theory-based analysis of search engine non-neutral behavior

Luis Guijarro, Vicent Pla, Bruno Tuffin, Patrick Maillé, Pierre Coucheney

To cite this version:

HAL Id: hal-00725180
https://hal.archives-ouvertes.fr/hal-00725180
Submitted on 24 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A game theory-based analysis of search engine non-neutral behavior

Luis Guijarro*, Vicent Pla*, Bruno Tuffin†, Patrick Maillé‡ and Pierre Coucheney‡

*Universitat Politècnica de València, Spain
†INRIA Rennes, France
‡Telecom Bretagne, France

Abstract—In recent years, there has been a rising concern about the policy of major search engines, and more specifically about their ranking in so-called organic results corresponding to keywords searches. The associated proposition is that their behavior should be regulated. The concern comes from search bias, which refers to search rankings based on some principle different from the expected automated relevance. In this paper, we analyze one behavior that results in search bias: the payment by content providers to the search engine in order to improve the chances to be located and accessed by a search engine user. A simple game theory-based model is presented where both a search engine and a content provider interact strategically, and the aggregated behavior of users is modeled by a demand function. The utility of each stakeholder when the search engine is engaged in such a non-neutral behavior is compared with the neutral case when no such side payment is present.

Index Terms—Search engine, neutrality, Nash equilibrium, user welfare

I. INTRODUCTION

In recent years there has been an increasing concern about Google policy, in terms of its search activities, implying the eventuality that this role should be regulated by government authorities. This is the aim of “search engine neutrality” advocates. The concern is about search engine bias, a term used to describe the activities of a search engine exercising its editorial discretion in a manner that advantages its own or affiliated content, which could favor some content wishing to pay to be better ranked, or which could disadvantage rivals. Search bias therefore refers to rankings based on some principle other than automated relevance for users. Within this paper, we focus on this concern shown by some regulators on those search rankings that benefit the host search engine: for example the European Commission justifies its inquiry into Google by a need to investigate if there is a conduct potentially “lowering the ranking of unpaid results” relatively to paid advertisements even in the regular, also called organic, results, as opposed to the sponsored links clearly declared (and presented as such to users) to make money [1].

Search neutrality advocates need, however, to demonstrate that there is a problem necessitating any of the various prescribed remedies, such as the application of standard merger analysis under the antitrust laws and even the creation of a government-sponsored “public option” for searches [2]. Problems cited by search neutrality advocates may be broadly classified in two groups: competition law and antitrust problems arising from “non-objective” search results, and social and cultural issues arising from the use that consumers may make with “non-objective” results. This paper is focused on the former group.

As regards the potential competitive harm from search bias, one argument is based on the “essential facilities” doctrine [3] and stresses that popular search engines, Google for instance, act as a “bottleneck” to access of websites to consumers. By using its power to determine which end websites are reached or not, a search engine can effectively exclude nascent websites from both advertising and sales revenue. Another, but related, argument claims that Google disadvantages its content rivals by raising their (awareness) costs relatively to its own. By directing search traffic to its own products — e.g., its mail, calendar, and marketplace platforms — Google would effectively discriminate against rivals and forces those rivals into more expensive substitute distributions channels [4]. The question remains whether a search engine’s use of its search algorithm to direct traffic to itself harms competition and consumers [5].

This paper aims to shed light on the issue of the search bias. Specifically, it focuses on the analysis of the harm that an apparently non-neutral behavior by a search engine may have on content providers and users. The non-neutral behavior we specifically consider here consists in charging a side payment to the content providers in exchange for better search results. While the issue of network neutrality, that is, the question whether all packets should be treated equally or not, has recently been the topic of a very active debate and extensive literature (see references [6], [7]), this paper is to our knowledge the first one dealing with a mathematical modeling and analysis of search neutrality, though this issue may be critical as we have just highlighted above.

The paper is a result of the project INNIS†. The project is run under the umbrella of the EuroNF Network of Excellence, which is financed through the FP7 program of the European Union.

†Visit the project webpage at http://www.irisa.fr/dionysos/pages_perso/tuffin/INNIS/
The paper is structured as follows. The next section describes the scenario to be modeled and presents the details of the baseline (neutral) model, which comprises a search engine without a search bias, and analyzes the corresponding pricing optimization problem of a pay-per-use content provider. Section III models the case where the search engine exhibits a search bias, and computes the related equilibrium of the non-cooperative game played with the content provider. Section IV compares the outputs obtained in both cases, neutral and non-neutral. And finally, some conclusions are drawn in Section V.

II. GENERAL MODEL, AND ANALYSIS OF THE NEUTRAL CASE

The scenario modeled in this paper is shown in Fig. 1. It includes:

- The users, who typically access the content stored at the content provider by using the search engine services by means of keywords.
- One content provider (CP), which provides paid content to the users, at a price \( p \).
- The search engine (SE), which helps the users in locating the content at the content provider, and in the non-neutral case charges a price \( q \) to the CP.

Arrows represent money flows, and are labeled by the corresponding unit price.

In this section, we model the case where there is no side payment from the CP to the SE, that is, where \( q = 0 \) in Fig. 1. We take this case as the baseline model, so that the search engine non-neutral behavior will be compared with this benchmark/baseline model.

A. Users

We model the users by means of a demand function \( D(p) \), representing the total amount of users subscribing to the CP, where \( p \) is the flat-rate price charged by the content provider. \( D \) is assumed linearly decreasing in \( p \), i.e.,

\[
D(p) = \beta_0 (D_0 - d \cdot p),
\]

where \( d > 0 \) is a parameter representing the sensitivity of users to the price \( p \), and \( \beta_0 \) denotes the probability \((0 < \beta_0 < 1)\) that the content is located by the search engine and therefore accessed by a user. \( D_0 \) is the maximum potential level of demand, if the content was for free, and provided it is fully advertised by the search engine. \( D_0 - dp \) can be interpreted as (being proportional to) the probability that a user, given the fact that he has “found” the content, actually subscribes to its service. That decision depends on the price \( p \) set by the content provider, but also on the user’s willingness-to-pay for the service.

B. Content provider

We assume as described in Fig. 1 that the CP charges a fixed price \( p \) to the users. In the baseline model the CP is assumed not to incur costs, so that the CP’s profits are given by:

\[
\Pi_c = D \cdot p.
\]

Again, we also assume that every piece of content that the users would like to access is stored by the content provider. However, the content will be located by the users by using the service of the search engine, which means that there is content stored at the CP which may end up not being accessed by the users if the search engine denies mentioning it.

C. Search engine

The role of the SE is to “help” users reach the content that is relevant to them.

We assume in our baseline model that the SE does not charge any usage-based price to the CPs. However, there is an additional revenue coming from sponsored links (displayed usually at the top and/or at the right of the regular/organic links), and which can reasonably be assumed to be proportional to the demand level. Indeed, users subscribing to the CP can be seen from the SE point of view as “satisfied” by the SE results, hence more likely to use that SE again and yield revenues through the sponsored links. Thus,

\[
\Pi_{SE} = \lambda \cdot D
\]

with \( \lambda > 0 \) the coefficient giving the revenue per subscriber to the CP service, coming from sponsored links. Therefore in the baseline model more demand at the CP level will give more revenue to the SE.

D. Solution

Given that the parameters \( D_0, \beta_0, d \) and \( \lambda \) are fixed, the only decision variable in this baseline model is the CP price \( p \). The CP will charge a price \( p \) so as to maximize its profits \( \Pi_c \). It is straightforward to see that this maximum is reached when the first order condition (FOC)

\[
\frac{\partial \Pi_c}{\partial p} = 0
\]

is fulfilled, giving

\[
p^{(n)} = \frac{D_0}{2d}.
\]

This then gives:

\[
D^{(n)} = \beta_0 \frac{D_0}{2},
\]

\[
\Pi_c^{(n)} = \beta_0 \frac{D_0^2}{4d},
\]

\[
\Pi_{SE}^{(n)} = \lambda \beta_0 \frac{D_0}{2}.
\]
III. NON-NEUTRAL MODEL

The neutral case being solved, we now focus on the non-neutral behavior of a SE implementing a side payment \( q \) from the CP to the SE. The effect of the side payment is two-fold:

- Paying \( q \) to the SE will increase the chances that the content is located and accessed by the users, when users decide to use the SE. We model this effect through an increasing function \( \beta_1(q) \).
- The more the SE charges to the CP, the less the users trust the search results because the search bias will increase and the reputation of the SE will decrease, and the less likely they will use the SE. We model this reputation effect through a decreasing function \( \beta_2(q) \).

As a consequence, the overall probability \( \beta_0 \) that the CP is seen by a user in the baseline model can be replaced here by a probability

\[
\beta(q) = \beta_1(q) \beta_2(q),
\]

taking into account both positive and negative consequences of side payments. We will also assume

\[
\beta(0) = \beta_0
\]

so that the CP not paying any charge to the SE will reasonably be such that the model comes down to the baseline/neutral case. In the non-neutral case though, the CP could play with \( \beta_1 \) to impose \( \beta(0) < \beta_0 \) to enforce the CP to pay. This (other) case is left for further research; but does not add any major change to the analysis.

Taking into account this side payment, the demand function \( D \) and the profits \( \Pi_c \) and \( \Pi_{SE} \) are now given by:

\[
D(p,q) = \beta(q)(D_0 - d \cdot p)
\]

\[
\Pi_c = D \cdot (p - q)
\]

\[
\Pi_{SE} = D_q + \lambda D_q
\]

A. General solution

In the non-neutral model, given \( \beta(\cdot) \) and parameters \( D_0, d, \) and \( \lambda, \) the CP and the SE interact strategically and non-cooperatively in order to maximize their respective profits. That is, the CP will charge a price \( p \) so as to maximize its profit \( \Pi_c \), and simultaneously the SE will charge a side payment \( q \) so as to maximize its profits \( \Pi_{SE} \). The analysis framework is that of non-cooperative game theory [8], and the equilibrium concept that of Nash equilibrium, which is a strategy profile \((p^*, q^*)\) such that no player (the CP or the SE) can unilaterally increase its profits through a price change.

Given the strategy \( q \) of the SE, the FOC for the CP is now:

\[
\frac{\partial \Pi_c}{\partial p} = \beta(q)(D_0 - 2dp + dq) = 0
\]

which, assuming \( \beta(q) \neq 0 \), yields

\[
p^{(nn)} = \frac{D_0}{2d} + q/2.
\]

And for the SE, differentiating with respect to \( q \) and equating to zero, we get

\[
\frac{\partial \Pi_{SE}}{\partial q} = (D_0 - dp) \left( (\gamma + \lambda) \frac{\partial \beta(q)}{\partial q} + \beta(q) \cdot 1 \right) = 0
\]

where we assume that \((D_0 - dp) > 0 \) (otherwise the SE gets nothing). As a consequence, the derivative is of the same sign as \( \beta(q) + \frac{\partial \beta(q)}{\partial q} + \lambda \). But remark that if

\[
\beta(q) + \frac{\partial \beta(q)}{\partial q} + \lambda = \begin{cases} 
0 & \text{if } 0 < q < q^*, \\
0 & \text{if } q = q^*, \\
> 0 & \text{if } q > q^*,
\end{cases}
\]

then the value \( q^* \) (provided the corresponding value \( p^{(nn)} \) in (15) leads to a positive demand) maximizes \( \Pi_{SE}(q) \). This is assumed from now on.

Additionally, the equilibrium \( q^* > 0 \) should lead to a better performance, in terms of probability for the CP to be visited, with respect to the neutral case, so that the non-neutral service is an enhanced service for the CP. That is,

\[
\beta(q^*) \geq \beta(0).
\]

B. Particular solution

We consider, in order to make some numerical computations, the following expressions for \( \beta_1(q) \) and \( \beta_2(q) \):

\[
\beta_1(q) = \frac{1 - \frac{1 - \beta_0}{q + 1}}{q + 1}
\]

\[
\beta_2(q) = \frac{1}{q + 1}.
\]

Note that the condition in (10) is satisfied. Remark that the constraint \( \beta(q^*) > \beta(0) \) and the fact that \( \beta_1(q)\beta_2(q) \) is decreasing when \( q > 1 - 2\beta_0 \) implies that \( \beta_0 < 1/2 \).

Substituting (19) and (20) in (16), we get

\[
q^* = \frac{\beta_0 + \lambda(1 - 2\beta_0)}{\lambda - 2 + \beta_0}.
\]

Requirement (18) becomes

\[
\beta(q^*) = \frac{(\lambda - 2 + \beta_0)(\lambda - \lambda\beta_0 - \beta_0 + \beta_0^2)}{4(\lambda - 2 + \beta_0)(\lambda - 1)^2} > \beta_0
\]

After some algebra and bearing in mind that \( \beta_0 < 1/2 \), the above inequality can be rewritten as

\[
\lambda^2 - 2\lambda - \frac{\beta_0(2 - 3\beta_0)}{(1 - 2\beta_0)^2} > 0.
\]

which is satisfied if and only if

\[
\lambda > 1 + \sqrt{1 + \frac{\beta_0(2 - 3\beta_0)}{(1 - 2\beta_0)^2}} = 2 + \frac{\beta_0}{1 - 2\beta_0}.
\]
Having justified that this $\beta(q)$ complies with the required properties, from (21) and (15), the following values can be obtained:

\[ D^{(nn)} = \beta(q^*)(D_0 - d \cdot p^{(nn)}) \]  \hspace{1cm} (23)

\[ \Pi_c^{(nn)} = D \cdot (p^{(nn)} - q^*) \]  \hspace{1cm} (24)

\[ \Pi_{SE}^{(nn)} = Dq^* + \lambda D. \]  \hspace{1cm} (25)

IV. COMPARISON BETWEEN THE NEUTRAL AND NON-NEUTRAL CASES

In this section, the output for the users, the CP and the SE in the non-neutral case is compared with the baseline case, and constraints on the values of the parameters providing better output thanks to non-neutrality are derived.

A. Demand and profits

We first focus on the users demand $D$, the CP profit $\Pi_c$, and the SE profit $\Pi_{SE}$.

As a first approximation, note that $p^{(nn)} = D_0/2d + q/2 \geq D_0/2d = p^{(n)}$. Thus, the price paid by users in the non-neutral case is higher that in the neutral case. However, the performance of the search engine increases at $q^*$ with respect to the situation at $q = 0$, so that it may happen that this effect dominates and the demand increases.

To simplify notation in the sequel we write $\beta^*$ instead of $\beta(q^*)$. We assume that (18) is satisfied, then $\beta_0/\beta^* < 1$. Dealing with (6)–(8) and with (23)–(25), the following conditions are obtained:

\[ D^{(nn)} > D^{(n)} \quad \text{iff} \quad C1 \equiv \frac{D_0}{d} > \frac{q^*}{1 - \frac{\lambda}{2\beta^*}}. \]  \hspace{1cm} (26)

\[ \Pi_c^{(nn)} > \Pi_c^{(n)} \quad \text{iff} \quad C2 \equiv \frac{D_0}{d} > \frac{q^*}{1 - \sqrt{\frac{\lambda}{2\beta^*}}}, \]  \hspace{1cm} (27)

\[ \Pi_{SE}^{(nn)} > \Pi_{SE}^{(n)} \quad \text{iff} \quad C3 \equiv \frac{D_0}{d} > \frac{q^*}{1 - \frac{x+y}{\lambda+y+2\beta^*}}. \]  \hspace{1cm} (28)

Comparing $C1$ and $C3$ (or directly (3) and (13)) it is easily seen that if $D^{(nn)} > D^{(n)}$ then necessarily $\Pi^{(nn)}_{SE} > \Pi^{(n)}_{SE}$. Similarly, if $\Pi^{(nn)}_{c} > \Pi^{(n)}_{c}$ then $D^{(nn)} > D^{(n)}$. In summary, we have that

\[ \Pi^{(nn)}_{c} > \Pi^{(n)}_{c} \implies D^{(nn)} > D^{(n)} \implies \Pi^{(nn)}_{SE} > \Pi^{(n)}_{SE}. \]  \hspace{1cm} (29)

Hence, the necessary and sufficient condition for both profits being higher in the non-neutral scenario is $C2$.

On the minimum value of $\lambda$: In the following paragraphs, we derive the restrictions that $C2$ imposes on $\lambda$.

Note that the right-hand side of the inequality in $C2$ is decreasing in $\lambda$. This follows from the two facts:

- $\beta^*$ is decreasing in $\lambda$

\[ \frac{\partial \beta^*}{\partial \lambda} = \frac{(\lambda - 2 + \beta_0)(\lambda - \beta_0)}{4(1 - \beta_0)(\lambda - 1)^2} \]

- $q^*$ is decreasing in $\lambda$

\[ \frac{\partial q^*}{\partial \lambda} = \frac{-2(1 - \beta_0)^2}{(\lambda - 2 + \beta_0)^2} < 0. \]

Therefore, for a given value of $D_0/d$, condition $C2$ is equivalent to imposing a minimum value for $\lambda$. Furthermore, the lower $D_0/d$ is, the higher that minimum value will be. Conversely, for a given $\lambda$, $C2$ gives a lower bound for $D_0/d$, and the higher $\lambda$ is, the less restrictive that lower bound is.

B. User welfare

Another output of the equilibrium to compare is the user welfare representing the aggregated value that users get from the whole service.

One expression for the user welfare can be obtained from the demand function $D(p, q) = \beta(q)/(D_0 - dp)$.

We have assumed that users differ not only with respect to their search behavior—modeled through a distribution on the SE choices and on the click behavior on search results over the user population, but also with respect to their willingness-to-pay—modeled using a distribution over the population.

For a given value of $q$, based on the form of the demand, the fraction of users willing to pay at least $p$ for the service, among those who found the CP, equals $1 - dp/D_0$ (assuming that all users would subscribe if $p = 0$). Varying $p$, the distribution of the willingness-to-pay among users is deduced to be uniform on $[0, D_0/d]$, while $D_0$ represents the total size of the user population.

A reasonable assumption, implicit in the product form of the demand function, is that the search behavior of users is independent of their willingness-to-pay for the service, i.e., the distribution of willingness-to-pay is independent of $q$. Consequently, for given values of $q$ and $p$ we can simply compute the (average) benefit that users make by finding and subscribing to the service offered by the CP, as the product of $D_0$ times $\beta(q)$ times the average user benefit from using the service given its price $p$.

This gives, recalling that only users with willingness-to-pay $w > p$ will subscribe:

\[ UW(p, q) = D_0 \beta(q) \int_{w = 0}^{D_0/d} d/D_0[w - p]^+ dw \]

\[ = \beta(q) d \int_{w = p}^{D_0/d} (w - p)^+ dw \]

\[ = \beta(q) d \int_{y = 0}^{D_0/d - p} y dy \]

\[ = \frac{1}{2} \beta(q) d (D_0/d - p)^2 \]

\[ = \frac{1}{2d} \frac{D^2(p, q)}{\beta(q)}. \]

where $x^+ = \max(0, x)$.

The above expression is applicable to both the neutral and
non-neutral cases, that is:

\[
\begin{align*}
\text{UW}^{(n)} &= \text{UW}(p^{(n)}, 0) = \frac{\beta_0}{8d} D_0^2 \\
\text{UW}^{(nn)} &= \text{UW}(p^{(nn)}, q^*) = \frac{\beta_0}{8d} (D_0 - dq^*)^2.
\end{align*}
\]

It is easy to check that

\[
\text{UW}^{(nn)} > \text{UW}^{(n)} \iff C4 \equiv \frac{D_0}{d} > \frac{q^*}{1 - \sqrt{\frac{\beta_0}{8d}}}
\]

which happens to be the same condition as in (27), and thus

\[
\text{UW}^{(nn)} > \text{UW}^{(n)} \iff \Pi^{(nn)}_c > \Pi^{(n)}_c \\
\Rightarrow D^{(nn)} > D^{(n)} \Rightarrow \Pi^{(nn)}_SE > \Pi^{(n)}_SE .
\]

C. Discussion

Let us now illustrate the results we have just derived analytically. Our goal is to discuss the influence that the parameters have on the outcome for each stakeholder.

a) On the conditions C2, C3 and C4: In Fig. 2, the domains such that conditions C2 and C3 hold are represented as a function of \(\lambda\).

First, note that the domain for C1 is not represented because, as derived in (30), \(C4 \Rightarrow C1\), and because the user welfare is more appropriate than the user demand when assessing whether the users are better off in neutral or non-neutral situations. Second, C4 is not represented since it is equivalent to C2. Third, the graph represents the right-hand side of each inequality as a function of the parameter \(\lambda\), and each pair of curves is generated with a different value for the parameter \(\beta_0\). And fourth, the represented range of \(\lambda\) complies with (22).

Note that \(C2 = C4\) is indeed the most constraining condition (as proved previously), since the threshold curve lies above that of C3 for every value of \(\beta_0\). Also the graph demonstrates that there is a wide range of values for \(D_0/d\) and \(\lambda\) such that C3 holds but C2 = C4 does not, i.e., where the SE is better off but both the CP and the users are harmed with a non-neutral SE.

b) On the values of \(\Pi_c\), \(\Pi_{SE}\) and UW: In Fig. 3, the values of \(\Pi_c\), \(\Pi_{SE}\) and UW are represented as a function of \(\beta_0\). The values of the parameters are \(D_0 = 30, d = 5\) and \(\lambda = 8\).

As expected from the conditions C2, C3 and C4 and the discussion of Fig. 2, for low values of \(\beta_0\), all stakeholders are better off in the non-neutral case. As \(\beta_0\) increases, the first stakeholders to be harmed are the users and the CP, while the SE remains as the only stakeholder to take profit from the search bias for high values of \(\beta_0\).

V. CONCLUSIONS AND FUTURE WORK

We conclude that under certain conditions, a side payment by the content provider to the search engine is beneficial for all stakeholders, since:

- the users are better off, which means that the improvement achieved by the increase in the likelihood to find the content compensates for the increase in the content price: more users “see” the content, and despite the fact that the price increases, the number of subscribers increases;
- the content provider benefits from a better visibility, allowing it to increase the subscription price to cover the payments to the search engine;
- and the search engine benefits from the increase in either the demand or the side payment, or in both.

Note that this win-win situation depends on the characteristics of the user demand, through the maximum tolerable price over the population; on the quality of the search activity; and on fraction of SE profits that depends on the users demand but
is independent of the side payment. As we have seen, for this situation to occur, it is required that:

- users be willing to pay a sufficient amount for the service,
- without side payments, the SE gets a sufficient revenue out of user demand for content,
- the visibility of the content in the neutral situation is low enough.

When this win-win situation is not achieved, the users and the CPs are the first stakeholders to suffer from non-neutral search activity.

The above results support in this case the demand of the search engine neutrality advocates, which claim that regulatory measures should be put in place so as to avoid that search bias results in the users being harmed.

We plan to extend the present work in order to tackle more complex and realistic scenarios. Specifically, we intent to extend the neutral and non-neutral model with two CPs, so that the following situations may be evaluated: (1) competition between the CPs; (2) collusion among the CPs; and (3) integration between the SE and one CP (such as Youtube owned by Google).

REFERENCES